Multilinear Discriminant Analysis through Tensor-Tensor Eigendecomposition

Randy C. Hoover
Department of Electrical and Computer Engineering
SD School of Mines and Technology
Rapid City, South Dakota
Email: randy.hoover@sdsmt.edu

Kyle Caudle
Department of Mathematics and Computer Science
SD School of Mines and Technology
Rapid City, South Dakota 57701
Email: kyle.caudle@sdsmt.edu

Karen Braman
Department of Mathematics and Computer Science
SD School of Mines and Technology
Rapid City, South Dakota 57701
Email: karen.braman@sdsmt.edu

***Abstract*—The current paper presents a new approach to dimensionality reduction and supervised learning for classification of multi-class data. The approach is based upon recent developments in tensor decompositions and a newly defined algebra of circulants. In particular, it is shown that under the right tensor multiplication operator, a third order tensor can be written as a product of third order tensors that is analogous to a traditional matrix eigenvalue decomposition where the "eigenvectors" become eigenmatrices and the "eigenvalues" become eigen-tuples. This new development allows for a proper tensor eigenvalue decomposition to be defined and has natural extension to tensor linear discriminant analysis (LDA). Comparisons are made with traditional LDA and it is shown that the current approach is capable of improved classification results for benchmark datasets involving faces, objects, and hand written digits.**

I. Introduction

Dimensionality reduction and supervised learning for classification of multi-class data is an important issue in many areas of machine learning, data analytics, and artificial intelligence. Subspace methods have been successfully applied to many problems relating to dimensionality reduction and classification in the imaging sciences. Specific examples include face characterization and recognition [1]–[5], object recognition and pose estimation [6]–[14], as well as a host of applications which arise in industrial automation [15].

Subspace methods take advantage of the fact that a set of highly correlated images can be represented by a low-dimensional subspace [16]. Arguably the most popular linear subspace methods in use are principal component analysis (PCA), linear discriminant analysis (LDA), and locally preserving projections (LPP) [1]–[4], [7], [16]–[19]. Traditionally, the reduced dimensional subspace of a large data-set has been computed using a linear algebraic framework. In particular, given a set of n images $\{I_1, I_2 \ldots, I_q\}$, each image is "row-scanned" into a column vector $\mathbf{x}$ of size $m = hv$. The collection of image vectors are then concatenated to form an "image data matrix" $X = [\mathbf{x}_1, \mathbf{x}_2 \cdots, \mathbf{x}_q] \in \mathbb{R}^{m \times q}$ where in general $m >> q$. Once the data matrix X of a set of samples has been constructed, any number of linear subspace techniques can be applied for dimensionality reduction and subsequent classification.

Although PCA, LDA, and LPP have been traditionally approached via linear algebraic methods, the natural representation of each image is lost during the "row-scanning" process, thereby eliminating the spatial correlation in each image. A more natural representation for the collection of images is to leave each image in matrix form, and stack each matrix into a tensor structure (also referred to as an n-way or n-mode array). As illustrated in [20], the tensor structure immediately lends itself to multi-linear algebraic operations on high-order tensors. Although there is no *unique* eigenvalue decomposition for high-order tensors, several tensor-singular value decompositions (SVD) have been proposed that attempt to mimic similar properties of the matrix SVD [14], [20]–[34]. While the N-mode SVD outlined in [20] has shown promising results for tensor PCA and ICA, the decomposition is constructed from products of matrices with a core-tensor as opposed to a product of tensors as one might expect.

In the current paper, we describe a new approach to supervised learning by extending traditional LDA to a multilinear framework (referred to as multilinear discriminant analysis (MLDA)). The approach is based on recent developments based upon Fourier theory and an algebra of circulants as outlined in [35]–[39]. It is shown that under the right tensor multiplication operator, a third order tensor can be written as a product of third order tensors in which the left tensor is a collection of eigenmatrices, the middle tensor is a front-face diagonal (denoted as f-diagonal) tensor of eigen-tuples, and the right tensor is the tensor inverse of the eigenmatrices resulting in a tensor-tensor eignevalue decomposition that is similar to its matrix counterpart. Experimental results are presented on three different benchmark data sets to include the MNIST data set of numerical hand written digits, the Ray-Traced data set of objects under different orientations, and the extended Yale B data set of human faces under different illumination conditions.

The remainder of this paper is organized as follows: In Section II we discuss the relevant tensor algebra, the newly defined tensor multiplication operator, and the tensor-tensor eigenvalue decomposition. In Section III we present an analysis of how the proposed tensor-tensor eigenvalue decomposition can be utilized to extend traditional LDA to a multilinear discriminant

framework. Section IV provides an overview of the different data sets used for validation of our proposed research as well as classification accuracy for each data set. Finally, Section V presents some discussion and provides some insight into future research directions.

II. Mathematical Foundations of Tensors

In the current section we discuss the mathematical foundations of the tensor decompositions used in the current work. While most of the theory in this section is outlined in [27], [35]–[37], we summarize this theory here to keep the current work self contained.

A. Mathematical Preliminaries

The term *tensor*, as used in the context of this paper, refers to a multi-dimensional array of numbers, sometimes called an *n-way* or *n-mode* array. If, for example, $\mathcal{A} \in \mathbb{R}^{\ell \times m \times n}$ then we say $\mathcal{A}$ is a third-order tensor where *order* is the number of ways or modes of the tensor. Thus, matrices and vectors are second-order and first-order tensors, respectively. Fundamental to the results presented in this paper is a recently defined multiplication operation on third-order tensors which itself produces a third-order tensor [35], [36].

Further, it has been shown in [37] that under this multiplication operation, $\mathbb{R}^{\ell \times m \times n}$ is a free module over a commutative ring with unity where the "scalars" are $\mathbb{R}^{1 \times 1 \times n}$ tuples. In addition, it has been shown in [37] and [38] that all linear transformations on the space $\mathbb{R}^{\ell \times m \times n}$ can be represented by multiplication by a third-order tensor. Thus, even though $\mathbb{R}^{\ell \times m \times n}$ is not strictly a vector space, many of the familiar tools of matrix linear algebra can be applied in this new context, including the basic building blocks of principal component analysis. For a more in depth discussion on this topic, the reader is referred to [37].

First, we review the basic definitions from [36] and [35] and introduce some basic notation. It will be convenient to break a tensor $\mathcal{A}$ in $\mathbb{R}^{\ell \times m \times n}$ up into various slices and tubal elements, and to have an indexing on those. The i^{th} lateral slice will be denoted $\mathcal{A}_i$ whereas the j^{th} frontal slice will be denoted $\mathcal{A}^{(j)}$. In terms of MATLAB indexing notation, this means $\mathcal{A}_i \equiv \mathcal{A}(:, i, :)$ while $\mathcal{A}^{(j)} \equiv \mathcal{A}(:, :, j)$.

We use the notation $\mathbf{a}_{ik}$ to denote the i, k^{th} tube in $\mathcal{A}$; that is $\mathbf{a}_{ik} = \mathcal{A}(i, k, :)$. The j^{th} entry in that tube is $\mathbf{a}_{ik}^{(j)}$. Indeed, these tubes have special meaning for us in the present work, as they will play a role similar to scalars in $\mathbb{R}$. Thus, we make the following definition:

Definition 1. *An element $\mathbf{c} \in \mathbb{R}^{1 \times 1 \times n}$ is called a* **tubal-scalar** *of length n.*

As mentioned previously, the set of tubal-scalars with length n endowed with element-wise addition and tensor multiplication (defined by the t-product in Def. 2) forms a commutative ring [37]. For ease of notation, we will use $\mathbf{0}$ to denote the additive identity, i.e., the tubal-scalar with all zero elements. Let $\mathbf{e}_1$ denote the tubal-scalar with all zero elements except a 1 in the first position. Then it is easy to see that $\mathbf{e}_1$ is the multiplicative identity in this ring and it will play an important role in the remaining tensor definitions.

In order to discuss multiplication between two tensors we must first introduce the concept of converting $\mathcal{A} \in \mathbb{R}^{\ell \times m \times n}$ into a block circulant matrix.

If $\mathcal{A} \in \mathbb{R}^{\ell \times m \times n}$ with $\ell \times m$ frontal slices then

$$\texttt{circ}(\mathcal{A}) = \begin{bmatrix} A^{(1)} & A^{(n)} & A^{(n-1)} & \dots & A^{(2)} \\ A^{(2)} & A^{(1)} & A^{(n)} & \dots & A^{(3)} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ A^{(n)} & A^{(n-1)} & \ddots & A^{(2)} & A^{(1)} \end{bmatrix},$$

is a block circulant matrix of size $\ell n \times mn$.

We anchor the `MatVec` command to the frontal slices of the tensor. $\texttt{MatVec}(\mathcal{A})$ takes an $\ell \times m \times n$ tensor and returns a block $\ell n \times m$ matrix

$$\texttt{MatVec}(\mathcal{A}) = \begin{bmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(n)} \end{bmatrix}.$$

The operation that takes $\texttt{MatVec}(\mathcal{A})$ back to tensor form is the `fold` command:

$$\texttt{fold}(\texttt{MatVec}(\mathcal{A})) = \mathcal{A}.$$

With these two operations in hand, we are in a position to introduce the t-product between two, third-order tensors that was developed in [35], [36]:

Definition 2. *Let $\mathcal{A} \in \mathbb{R}^{\ell \times p \times n}$ and $\mathcal{B} \in \mathbb{R}^{p \times m \times n}$ be two third order tensors. Then the* ***t-product*** *$\mathcal{A} * \mathcal{B} \in \mathbb{R}^{\ell \times m \times n}$ is defined as*

$$\mathcal{A} * \mathcal{B} = \texttt{fold}\left(\texttt{circ}(\mathcal{A}) \cdot \texttt{MatVec}(\mathcal{B})\right).$$

Note that in general, the t-product of two tensors will not commute. There is one special exception in which the t-product always commutes: the case when $\ell = p = m = 1$, that is, when the tensors are tubal-scalars.

Definition 3. *The* ***identity tensor*** *$\mathcal{I} \in \mathbb{R}^{m \times m \times n}$ is the tensor whose frontal slice is the $m \times m$ identity matrix, and whose other frontal slices are all zeros.*

Definition 4. *If $\mathcal{A}$ is $\ell \times m \times n$, then the* ***tensor transpose*** *$\mathcal{A}^T$ is the $m \times \ell \times n$ tensor obtained by transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through n.*

Definition 5. *A tensor $\mathcal{A} \in \mathbb{R}^{n \times n \times n}$ has an* ***tensor inverse*** *$\mathcal{B}$ provided*

$$\mathcal{A} * \mathcal{B} = \mathcal{I} \text{ and } \mathcal{B} * \mathcal{A} = \mathcal{I}$$

where $\mathcal{I} \in \mathbb{R}^{n \times n \times n}$. The tensor inverse is computed as

$$\mathcal{A}^{-1} = \texttt{fold}(\texttt{circ}(\mathcal{A})^{-1}).$$

Definition 6. *The tensor norm used throughout this paper is the Frobenious norm which for the tensor $\mathcal{A} \in \mathbb{R}^{\ell \times m \times n}$ is given by*

$$\|\mathcal{A}\|_F = \sqrt{\sum_{i=1}^{\ell}\sum_{j=1}^{m}\sum_{k=1}^{n} a_{ijk}^2}. \tag{1}$$

where a_{ijk} is the i, j, k^{th} element of $\mathcal{A}$.

B. Computation of the t-eigenvalue decomposition

The final tool necessary for a multilinear LDA is to define a tensor-tensor eigenvalue decomposition. In [28], [29], [37] the authors show that, for $\mathcal{A} \in \mathbb{R}^{n\times n\times n}$, there exists an $n \times n \times n$ tensor $\mathcal{P}$ and an $n \times n \times n$ f-diagonal tensor $\mathcal{D}$ such that

$$\mathcal{A} = \mathcal{P} * \mathcal{D} * \mathcal{P}^{-1} \implies \mathcal{A} * \mathcal{P} = \mathcal{D} * \mathcal{P} \implies \mathcal{A} * \mathcal{P}_j = \mathcal{P}_j \mathbf{d}_j \tag{2}$$

where $\mathcal{P}$ is an invertible tensor of eigenmatrices, and $\mathcal{D}$ is an f-diagonal tensor of eigentuples, denoted individually by $\mathbf{d}_j = \mathcal{D}(j, j, :)$. Throughout this paper, we refer to this decomposition as the **t-eig**. A graphical illustration of the **t-eig** is shown in Fig. 1.

Computation of the **t-eig** comes from the constructive proof outlined in [28], [29], [37], which will be restated here for completeness. It is well known in matrix theory that a circulant matrix can be diagonalized via left and right multiplication by a discrete Fourier transform (DFT) matrix. Similarly, a block circulant matrix can be block diagonalized via left and right multiplication by a block diagonal DFT matrix. For example, consider the tensor $\mathcal{A} \in \mathbb{R}^{n\times n\times n}$, then

$$(F_n \otimes I_n)\texttt{circ}(\mathcal{A})(F_n^* \otimes I_n) = \begin{bmatrix} D_1 & & & \\ & D_2 & & \\ & & \ddots & \\ & & & D_n \end{bmatrix},$$

where each of the D_i are $n \times n$, I is an $n \times n$ identity matrix of dimension, F_n is the $n \times n$ DFT matrix, F_n^* is its conjugate transpose, and $\otimes$ is the Kronecker product. To construct the **t-eig** defined in (2), the matrix eigenvalue decomposition is performed on each of the D_i, i.e., $D_i = P_i \Lambda_i P_i^{-1}$ resulting in the decomposition shown in (3). Applying $(F_n^* \otimes I_n)$ to the left and $(F_n \otimes I_n)$ to the right of each of the block diagonal matrices on the right hand side of (3) results in each being block circulant, i.e., if we define $\hat{P}$ as the block diagonal matrix with P_i as its diagonal blocks, then

$$(F_n^* \otimes I_\ell)\hat{P}(F_n \otimes I_\ell) = \begin{bmatrix} P_1 & P_n & \cdots & P_{n-1} \\ P_2 & P_1 & \cdots & P_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ P_n & P_{n-1} & \cdots & P_1 \end{bmatrix}.$$

Taking the first block column of each block circulant matrix and applying the `fold` operator results in the decomposition $\mathcal{P} * \mathcal{D} * \mathcal{P}^{-1}$. Note that for simplicity, as well as computational efficiency, this entire process can by performed using the fast Fourier transform in place of the DFT matrix as illustrated in [28], [29], [35]–[37].

III. Overview of Discriminant Analysis

A. Linear Discriminant Analysis

This section provides an overview of the traditional approach to LDA as outlined in [3], [16], [40]. The general idea behind LDA is to compute a projection matrix U that maximizes the between class means while simultaneously minimizing the within class means. Such a projection is computed by deriving two scatter matrices to account for variation both within and between different classes. We consider a set of data samples that contain C classes (with associated class labels) and each class contains N_i data samples, i.e., $i = 1, 2, \ldots, C$. Denoting each class as c_i we construct the within-class scatter matrix as

$$S_W = \sum_{i=1}^{C} S_i \tag{4}$$

where

$$S_i = \sum_{\mathbf{x}\in c_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \tag{5}$$

$\mathbf{x} \in \mathbb{R}^m$ is an m-dimensional data sample, and $\mathbf{m}_i$ is the mean of class i, i.e.,

$$\mathbf{m}_i = \frac{1}{N_i}\sum_{\mathbf{x}\in c_i} \mathbf{x}. \tag{6}$$

The between-class scatter matrix is constructed to account for the class means around the total mean of the data as

$$S_B = \sum_{i=1}^{C} N_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \tag{7}$$

where $\mathbf{m}$ is the mean of all data samples, and $\mathbf{m}_i$ is defined in (6). The projection matrix U can then be computed by maximizing the ratio of determinants between S_W and S_B in the projection space as

$$\arg\max_U \frac{|U^T S_B U|}{|U^T S_W U|}. \tag{8}$$

Re-casting (8) as a constrained optimization problem, it can be shown that the solution is computed by solving the generalized eigenvalue problem

$$S_B \mathbf{u}_p = \lambda_p S_W \mathbf{u}_p \tag{9}$$

where $U = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p]$ corresponds to the p largest eigenvalues λ_p [3], [16], [40]. Note that there are at most $C-1$ nonzero eigenvalues of (9) therefore the projection space has at most dimension $C - 1$.

B. Multilinear Discriminant Analysis

It is important to note that for imaging applications, traditional matrix LDA requires an image to be "row-scanned" into an image vector to construct the m-dimensional data sample $\mathbf{x}$. In particular, consider a square gray-scale image described by an $n \times n$ array of square pixels with intensity values normalized between 0 and 1. The image is represented by a matrix[1] $I \in [0, 1]^{n\times n}$ and the data sample would be

[1]Note that the image matrix I should not be confused with the standard identity matrix.

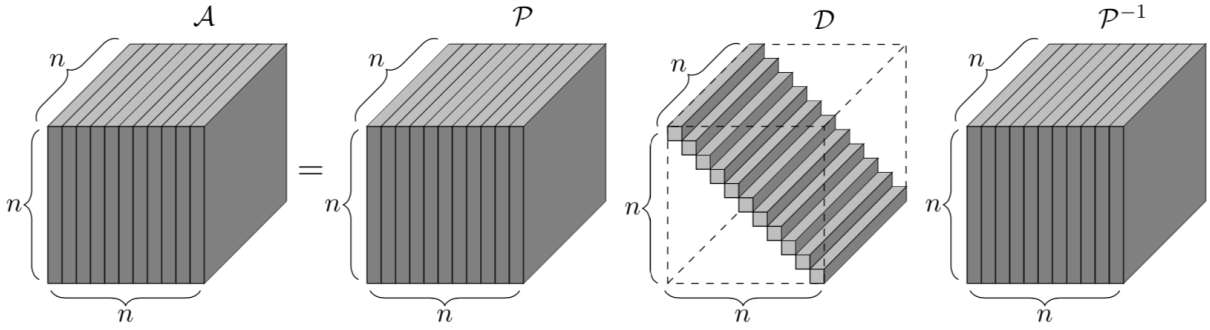

Fig. 1. Graphical illustration of the **t-eig** of an $n \times n \times n$ tensor.

$$\begin{bmatrix} D_1 & & \\ & \ddots & \\ & & D_n \end{bmatrix} = \begin{bmatrix} P_1 & & \\ & \ddots & \\ & & P_n \end{bmatrix} \begin{bmatrix} \Lambda_1 & & \\ & \ddots & \\ & & \Lambda_n \end{bmatrix} \begin{bmatrix} P_1^{-1} & & \\ & \ddots & \\ & & P_n^{-1} \end{bmatrix} \tag{3}$$

of dimension $m = n^2$. This poses some numerical issues when constructing the scatter matrices because, in general, the sample dimension (total number of pixels in the image) is generally much greater then the total number of samples. This is commonly referred to as the *small sample size* problem to which several regularized versions of LDA, or pre-projections (such as principal component analysis), have been developed to overcome such drawbacks.

An alternate approach to overcoming the *small sample size* problem encountered in imaging applications, is to keep the images themselves in "matrix" format to construct a high order tensor (cube of numbers) and perform LDA on the cube instead. Previous multilinear extensions have been proposed using either the Tucker, PARAFAC/CONDECOMP (CP), or HOSVD decomposition [21]–[24], [41], however all three methods rely on the product of matrices with a "core" tensor for the decompositions. In this section, we present an alternate approach to multilinear LDA by utilizing the definitions outlined in Section II along with the development of the **t-eig**, referred to as multilinear LDA (MLDA). Toward this end, we construct our data tensor $\mathcal{X} = [{}^1\mathcal{X}, {}^2\mathcal{X}, \cdots, {}^C\mathcal{X}] \in \mathbb{R}^{n \times q \times n}$ where each lateral slice of $\mathcal{X}$ is an $n \times n$ "sample" (image), ${}^i\mathcal{X}$ is a tensor corresponding to class i, q is the total number of samples, and again we assume C distinct classes[2]. The within-class and between-class scatter tensors can then be constructed as

$$\mathcal{S}_W = \sum_{i=1}^{C} \mathcal{S}_i \tag{10}$$

where

$$\mathcal{S}_i = \sum_{\mathcal{X}_j \in c_i} (\mathcal{X}_j - \mathcal{M}_i) * (\mathcal{X}_j - \mathcal{M}_i)^T \tag{11}$$

[2]We present the MLDA derivation assuming square images for simplicity, however, this is not a requirement in practice.

where we remind the reader that $\mathcal{X}_j \in \mathbb{R}^{n \times n}$ is the j^{th} lateral slice of $\mathcal{X}$, $\mathcal{M}_i \in \mathbb{R}^{n \times 1 \times n}$ is the mean tensor of class i, i.e.,

$$\mathcal{M}_i = \frac{1}{N_i} \sum_{\mathcal{X}_j \in c_i} \mathcal{X}_j, \tag{12}$$

and the transpose operator is that outlined in **definition 4**.

Similarly, the between-class scatter tensor is constructed as

$$\mathcal{S}_B = \sum_{i=1}^{C} N_i (\mathcal{M}_i - \mathcal{M}) * (\mathcal{M}_i - \mathcal{M})^T \tag{13}$$

where $\mathcal{M}$ is the mean of all data samples, $\mathcal{M}_i$ is defined in (12), and again the transpose operator is that outlined in **definition 4**.

The projection tensor $\mathcal{U}$ can then be computed by solving the generalized tensor eigenvalue problem

$$\mathcal{S}_B * \mathcal{U}_p = \boldsymbol{\lambda}_p * \mathcal{S}_W * \mathcal{U}_p \tag{14}$$

where $\mathcal{U} = [\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_p] \in \mathbb{R}^{n \times p \times n}$ are the eigenmatrices corresponding to the p largest eigentuples $\boldsymbol{\lambda}_p \in \mathbb{R}^{1 \times 1 \times n}$. Note that similar to its matrix counterpart, there are at most $C - 1$ nonzero eigentuples of (14) therefore the projection space has at most dimension $C - 1$. In addition, because the within- and between-class scatter tensors are of size $n \times n \times n$, the *small sample size* issue is non-existent and the computation of $\mathcal{U}$ can be performed via utilization of the **t-eig** as

$$\left(\mathcal{S}_W^{-1} * \mathcal{S}_B\right) * \mathcal{U}_p = \boldsymbol{\lambda}_p * \mathcal{U}_p \tag{15}$$

where $\mathcal{S}_W^{-1}$ is computed using **definition 5**.

IV. Data Sets and Experimental Validation

To validate the performance of our proposed MLDA algorithm as compared to traditional matrix LDA, we evaluated classification rates of three benchmark image databases. The first database is the MNIST database of handwritten characters [42], the second was a database of ray-traced CAD models under various pose [12], [13], [27], and the third was the extended Yale B database of human faces [43], [44].

A. MNIST

The MNIST data set consists of gray-scale images of hand written numeric digits in the range of 0 - 9. The data set was constructed by taking 30,000 digits from the original NIST Special Database 3 (SD-3) and 30,000 digits from the original NIST Special Database 1 (SD-1) to generate 60,000 training samples in total. The data contains samples from approximately 250 different writers collected among both Census Bureau employees (SD-3) and high-school students (SD-1) for the purpose of learning to accurately classify hand written numeric digits. An additional 5000 characters from both SD-1 and SD-3 were collected to be used as a testing set for evaluation purposes. While both the testing and training sets came from the same original databases, the collectors made sure the training and testing sets were disjoint to eliminate the possibility of contaminating classification results [42].

B. Ray-Traced Objects

The Ray-Traced object dataset considers image sequences where an "object" is placed at the center of an imaging sphere, as seen in Fig. 2, and a sample image is acquired at each of the black dots on the surface of the sphere. Setting β to a constant results in a one-dimensionally correlated sequence of images, ideally suited for classification and subsequent pose estimation of known rigid objects [12], [13], [27]. The data set consists of 128 images captured of 20 different objects with a single illumination source that is camera aligned for a total of 2560 training images. In addition, 64 images of each object were captured with the same co-latitudinal camera angle (β), however the object pose (α) was selected at random for a total of 1280 testing images. Similar to the MNIST dataset, the training and testing datasets were disjoint to eliminate the possibility of contaminating classification results.

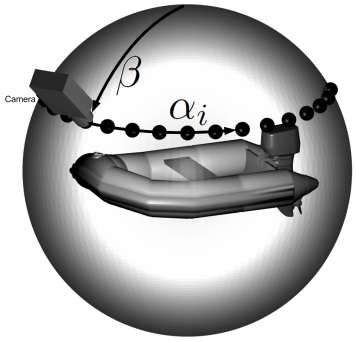

Fig. 2. This figure depicts sampling an object along a line of constant co-latitude, which results in a sequence of images correlated in a single dimension as outlined in [27]. A sample image is captured at each of the black dots that reside on the imaging sphere.

The images of each object in the training set were generated by ray-tracing CAD models provided by [45]. Similarly, the images captured for the testing set were also ray-traced from the same CAD models, however, zero mean Gaussian noise was added to each image in an effort to mimic images captured from a real-world camera. The data set is ideally suited for evaluating classification and pose estimation of objects that would be difficult to acquire large real-world image sets for training and evaluating learning algorithms.

C. Extended Yale B Face Database

The final data set evaluated in the current research is the extended Yale B data set. The extended Yale B dataset consists of images of 38 individuals from the frontal direction under 64 different illumination conditions. The subjects were placed at the center of a spherical cage where 64 of the frontal vertices of the cage contain an illumination source, and a frontal image was captured while each source was illuminated [43], [44]. In the original data set, 18 images of 7 of the subjects were corrupt, therefore, to keep the total number of samples consistent throughout, a random selection of 59 of the illumination directions per subject were used. In addition, unlike the previous two data sets, the extended Yale B data set does not contain both a training and testing set of images. Therefore, 40 of the images in the data set were used for training (1520 training samples), while the remaining 19 images (722 images) were used for testing.

D. Experimental Results

In each of the three data sets, each data sample is a square gray-scale image of size $n \times n$ and centered based on the specific data sample itself (i.e., MNIST is centered based on the center of mass of the digit whereas Ray-Traced is centered based on the center of the object of interest). A subset of example images for each of the three data sets are illustrated in Fig. 3 (a) MNIST, (b) Ray-Traced, and (c) extended Yale B. In the context of the current research, the construction of the data matrix X for LDA and the data tensor $\mathcal{X}$ for MLDA were performed as follows:

LDA:

1) Convert each sample to an image vector $\mathbf{x} \in \mathbb{R}^{n^2 \times 1}$ by "row-scanning" the original image.
2) Concatenate each training sample to construct the data matrix $X \in \mathbb{R}^{n^2 \times q}$, where the first N_1 columns of X correspond to samples drawn from class 1, the next N_2 columns are samples drawn from class 2, etc. and q corresponds to the total number of training samples.
3) Perform LDA on the matrix X to compute the project matrix $U \in \mathbb{R}^{n^2 \times (C-1)}$.
4) Project the data samples onto the low-dimensional subspace U, via $M = U^T X \in \mathbb{R}^{(C-1) \times q}$ for subsequent classification.

MLDA:

1) Convert each sample to a image tensor $\mathcal{X} \in \mathbb{R}^{n \times 1 \times n}$ by laterally rotating the original image.
2) Concatenate each training sample to construct the data tensor $\mathcal{X} \in \mathbb{R}^{n \times q \times n}$, where the first N_1 lateral slices of $\mathcal{X}$ correspond to samples drawn from class 1, the next N_2 lateral slices are samples drawn from class 2, etc.

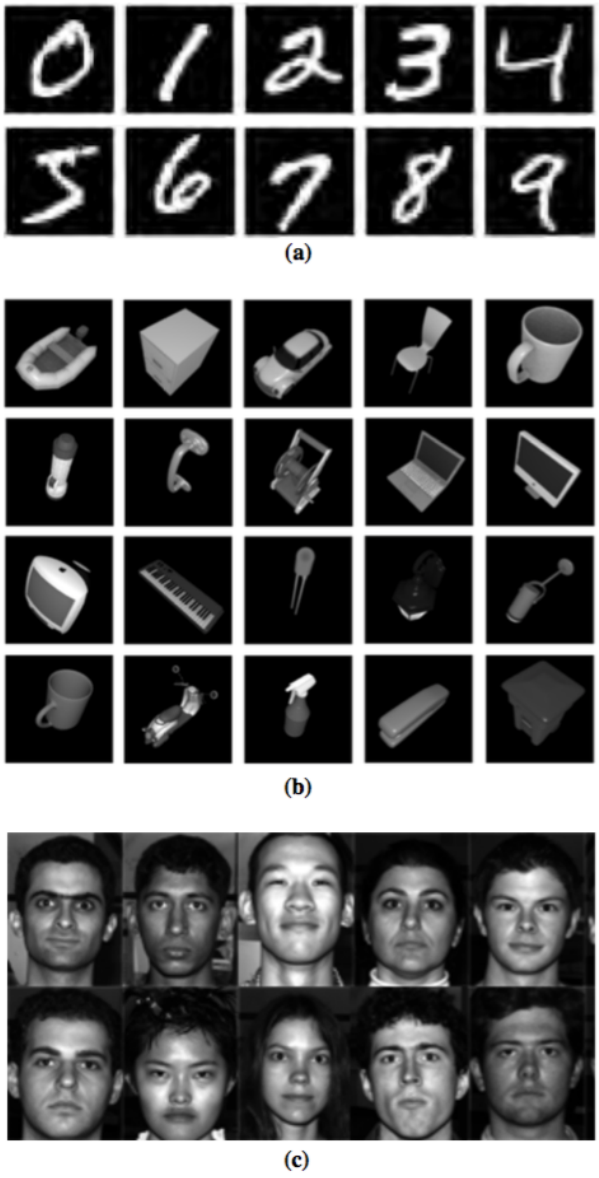

Fig. 3. A subset of images from each data set used in this research. (a) The MNIST data set, (b) images of Ray-traced CAD models courtesy of [45] sampled as outlined in [27], and (c) the extended Yale B face database.

3) Perform MLDA on the tensor $\mathcal{X}$ to compute the project tensor $\mathcal{U} \in \mathbb{R}^{n \times (C-1) \times n}$.
4) Project the data samples onto the low-dimensional subspace $\mathcal{U}$, via $\mathcal{M} = \mathcal{U}^T * \mathcal{X} \in \mathbb{R}^{(C-1) \times q \times (C-1)}$ for subsequent classification[3].

Once the projections M and $\mathcal{M}$ have been computed, classification is performed by projecting the test samples onto the corresponding subspace and performing both a nearest neighbor (NN) and nearest center (NC) search for the closest match. The details for each individual data set presented above, as well as the classification accuracy for each data set using both LDA and MLDA, are outlined in Table I. As can be seen from the table, our proposed MLDA approach provides better classification for all three of the databases tested (numerical digits, objects, and faces). Another interesting result is illustrated on the extended Yale B data set. Because the number of training samples is smaller than the other two data sets, and the image resolution is similar, the extended Yale B data set suffers from the *small sample size* problem outlined in Section III. Therefore, for completeness, a PCA pre-projection was also performed prior to an application of LDA (as outlined in [3]). These results are also present in Table I denoted by the (w/PCA). As can be seen in the table, using regularization significantly improves the classification rate. Moreover, the classification rate is much closer to that proposed by our approach that does not utilize any form of PCA pre-projection.

[3]The transpose and * here are the tensor-transpose and t-product respectively.

TABLE I
NUMERICAL DETAILS AND CLASSIFICATION ACCURACY FOR EACH OF THE THREE INDIVIDUAL DATA SETS OUTLINED IN SECTION IV.

	MNIST	Ray-Traced	extended Yale B
n	28	50	40
q	60,000	2560	1520
N_i	1000	128	40
C	10	20	38
# of Test Samples	10,000	1280	722
Method	MNIST	Ray-Traced	extended Yale B
LDA (NN):	85.1%	99.6%	56.8% (w/PCA: 97.8%)
LDA (NC):	84.7%	97.9%	55.2% (w/PCA: 96.2%)
MLDA (NN):	**93.2%**	**100%**	**95.7%**
MLDA (NC):	**91.4%**	**99.2%**	**94.8%**

V. CONCLUSIONS AND FUTURE DIRECTIONS

This paper presented a new approach to dimensionality reduction and subsequent classification of multi-class data using tensor decompositions and an algebra of circulants. In particular, it was shown that using a newly defined tensor multiplication operator, a third order tensor can be decomposed into the product of third order tensors, similar to a matrix eigenvalue decomposition. Using these results, a new approach to multilinear discriminant analysis was developed, referred to as MLDA. An analysis was presented in the context of classification of three different benchmark data sets, the MNIST data set of handwritten digits, the Ray-Traced data set of objects under different orientation, and the extended Yale B face data set consisting of frontal face images of individuals under different illumination conditions. It was illustrated that for all three data sets, our current approach outperformed traditional LDA in terms of classification accuracy. In addition, it was illustrated that our proposed MLDA performs similar to traditional LDA with a PCA pre-projection (i.e., our approach resolves the small sample size issue commonly associated with image classification using LDA). Future research directions are numerous, however most pressing will be investigating extension to other common dimensionality reduction techniques. Further, the authors would like to investigate nonlinear approaches to dimensionality reduction and classifica-

tion through a reproducing kernel Hilbert space (well known kernel-trick) applied to our multilinear algebraic framework.

ACKNOWLEDGMENT

This publication represents research supported by the NAVSEA Warfare Centers NEEC program, grant No. N00174-18-1-0001 and the NASA South Dakota Space Grant Consortium. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Naval Engineering Education Consortium.

REFERENCES

[1] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. PAMI*, vol. 12, no. 1, pp. 103–108, Jan. 1990.

[2] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, Mar. 1991.

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. PAMI*, vol. 19, no. 7, pp. 711–720, July 1997.

[4] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *IEEE Conf. Comp. Vis. and Patt. Rec.*, Seattle, WA, June 1994, pp. 84–91.

[5] M. H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Trans. PAMI*, vol. 24, no. 1, pp. 34–58, Jan. 2002.

[6] H. Murase and S. K. Nayar, "Illumination planning for object recognition using parametric eigenspaces," *IEEE Trans. PAMI*, vol. 16, no. 12, pp. 1219–1227, Dec. 1994.

[7] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance," *Int. J. Comp. Vis.*, vol. 14, no. 1, pp. 5–24, Jan. 1995.

[8] S. K. Nayar, S. A. Nene, and H. Murase, "Subspace methods for robot vision," *IEEE Trans. Robot. Automat.*, vol. 12, no. 5, pp. 750–758, Oct. 1996.

[9] C. Y. Chang, A. A. Maciejewski, and V. Balakrishnan, "Fast eigenspace decomposition of correlated images," *IEEE Trans. Image Proc.*, vol. 9, no. 11, pp. 1937–1949, Nov. 2000.

[10] R. C. Hoover, A. A. Maciejewski, and R. G. Roberts, "Pose detection of 3-D objects using S^2-correlated images and discrete spherical harmonic transforms," in *IEEE Int. Conf. Robot. Automat.*, Pasadena, CA, May. 2008, pp. 993–998.

[11] R. C. Hoover, A. A. Maciejewski, and R. G. Roberts, "Pose detection of 3-D objects using images sampled on $SO(3)$, spherical harmonics, and Wigner-D matrices," in *IEEE Conf. on Automat. Sci. and Engr.*, Washington DC, Aug 2008, pp. 47–52.

[12] R. C. Hoover, "Pose estimation of spherically correlated images using eigenspace decomposition in conjunction with spectral theory," Ph.D. dissertation, Colorado State University, 2009.

[13] R. C. Hoover, A. A. Maciejewski, and R. G. Roberts, "Eigendecomposition of images correlated on S^1, S^2, and $SO(3)$ using spectral theory," *IEEE Trans. Image Proc.*, vol. 18, no. 11, pp. 2562–2571, Nov. 2009.

[14] R. C. Hoover, A. A. Maciejewski, and R. G. Roberts, "Fast eigenspace decomposition of images of objects with variation in illumination and pose," *IEEE Tran. Sys. Man, Cyber. B: Cybernetics*, vol. PP, no. 99, pp. 1–12, Aug. 2010.

[15] E. N. Malamasa, E. G. Petrakisa, M. Zervakisa, L. Petit, and J.-D. Legat, "A survey on industrial vision systems, applications and tools," *Image and Vision Comp.*, vol. 21, no. 2, pp. 171–188, 2003.

[16] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. London, U.K.: Academic, 1990.

[17] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *J. Opt. Soc. Amer.*, vol. 4, no. 3, pp. 519–524, Mar. 1987.

[18] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. PAMI*, vol. 7, no. 3, pp. 328–340, Mar. 2005.

[19] D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal Laplacianfaces for face recognition," *IEEE Trans. Image Proc.*, vol. 15, no. 11, pp. 3608–3614, Nov. 2006.

[20] M. Alex, O. Vasilescu, and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *European Conf. on Comp. Vis.*, Copenhagen, Denmark, May 2002, pp. 447 – 460.

[21] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, Sept. 1966.

[22] R. A. Harshman, "Foundations of the PARFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis," University of California Los Angelas, Tech. Rep. 10,085, December 1970.

[23] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, March 2000.

[24] O. Vasilescu and D. Terzopoulos, "Multilinear projection for appearance-based recognition in the tensor framework," in *Int. Conf. on Comp. Vis.*, 2007, pp. 1–8.

[25] O. Vasilescu and D. Terzopoulos, "Multilinear subspace analysis of image ensembles," in *Int. Conf. on Comp. Vis. and Patt. Rec.*, 2003, pp. 93–99.

[26] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A survey of multilinear subspace learning for tensor data," *Pattern Recognition*, vol. 44, no. 7, pp. 1540–1551, July 2011.

[27] R. C. Hoover, K. S. Braman, and N. Hao, "Pose estimation from a single image using tensor decomposition and an algebra of circulants," in *Int. Conf. on Intel. Robots and Sys.*, 2011.

[28] N. Hao, M. E. Kilmer, K. S. Braman, and R. C. Hoover, "New tensor decompositions with applications in facial recognition," *SIAM Journal on Imaging Science (SIIMS)*, vol. 6, no. 1, pp. 437–463, Feb. 2013.

[29] M. E. Kilmer, K. S. Braman, N. Hao, and R. C. Hoover, "Third order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM Journal on Matrix Analysis and Applications (SIMAX)*, vol. 34, no. 1, pp. 148–172, Feb. 2013.

[30] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, Aug. 2009.

[31] P. M. Kroonberg, *Applied Multiway Data Analysis*. New York: Wiley, 2008.

[32] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, *Multilinear Subspace Learning*. Chapman & Hall/CRC, 2013.

[33] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang, "Multilinear discriminant analysis for face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 212–220, Jan 2007.

[34] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 103–123, Jan 2009.

[35] M. E. Kilmer, C. D. Martin, and L. Perrone, "A third-order generalization of the matrix SVD as a product of third-order tensors," Tufts University, Department of Computer Science, Tech. Rep. TR-2008-4, October 2008.

[36] M. E. Kilmer and C. D. Moravitz Martin, "Factorization strategies for third-order tensors," *Linear Algebra and Its Applications*, no. Special Issue in Honer of G.W.Stewart's 75^{th} birthday, 2009.

[37] K. Braman, "Third-order tensors as linear operators on a space of matrices," *Linear Algebra and its Applications*, vol. 433, no. 7, pp. 1241 – 1253, 2010.

[38] M. Kilmer, K. Braman, and N. Hao, "Third order tensors as operators on matrices: A theoretical and computational framework," Tufts University, Department of Computer Science, Tech. Rep. TR-2011-01, January 2011.

[39] D. F. Gleich, C. Greif, and J. M. Varah, "The power and arnoldi methods in an algebra of circulants," *arXiv*, vol. 1101.2173v1, 2011.

[40] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, pp. 179–188, 1936.

[41] O. Vasilescu and D. Terzopoulos, "A tensor algebraic approach to image synthesis, analysis and recognition," in *Int. Conf. on 3-D Dig. Imaging and Modeling*, 2007, pp. 3–12.

[42] (2018) MNIST database of handwritten digits. [Online]. Available: http://yann.lecun.com/exdb/mnist/

[43] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. PAMI*, vol. 23, pp. 643–660, June 2001.

[44] K. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. PAMI*, vol. 27, pp. 684–698, May 2005.

[45] (2007) Kator Legaz: 3-D model database for Blender. [Online]. Available: http://www.katorlegaz.com/